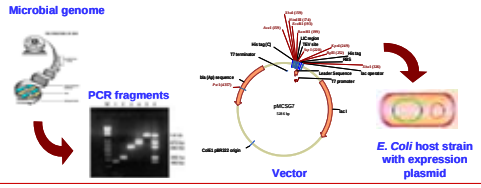
High Throughput Cloning for Protein Expression

What Do Cells Do With All Those Proteins?

Sarah Giuliani, Terese Peppler, Elizabeth Landorf, Frank Collart
Biosciences Division

Our high throughput procedure begins by examining the microbial genome to identify genes of interest. These genes are amplified from the genomic DNA by use of the polymerase chain reaction (PCR). The DNA segments are treated with the T4 polymerase enzyme and cloned into a special vector that allows for expression of the protein and replication of the plasmid in *Escherichia coli*. This vector is used to transform *E. coli* to generate a new strain that can used for expression of the target protein. We have developed a high throughput procedure to automate this process and enable the cloning of entire bacterial genomes.

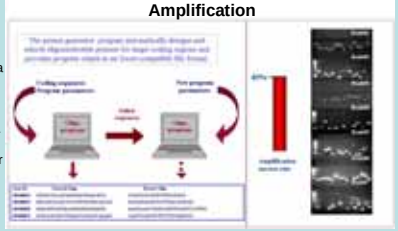


We developed a high throughput cloning and expression strategy from target to validated expression clone. This strategy evolved after evaluation of three critical elements common to many high throughput processes: targets, methods and screening requirements. Integration of these considerations into a series of methods results in an efficient process that has been scaled to generate thousands of *E. coli* clones. The pipeline incorporates automated liquid handlers that allow implementation of parallel processes. The process uses 96-well plates enabling the process of hundreds to thousands of targets in a single experiment.

Automated Cloning and Expression Process

High-throughput approaches for gene cloning and expression require the development of new non-standard tools for use by molecular biologists and biochemists. We have developed a web-based tool to design primers specifically for the generation of expression clones for both lab scale and high-throughput projects. The application allows the user complete flexibility to specify primer design parameters and minimizes the amount of manual intervention needed to generate large numbers of primers for simultaneous amplification of multiple target genes.

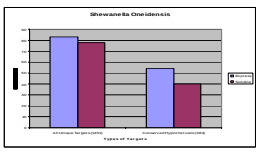
Oligonucleotide Design and PCR Amplification



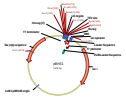
Expression and solubility analysis can be carried out on the automated system. Small scale expression cultures were grown at 37°C to early log phase and protein expression induced by the addition of isopropyl-β-galactose (IPTG) to 1mM. After solubility analysis proteins were grouped into four relative categories (insoluble; low, moderate, and high solubility) based on qualitative assessment of band intensity (Coomassie blue staining) after denaturing gel electrophoresis.

Expression Plate Layout and SDS-PAGE Gel

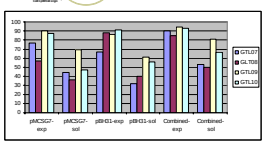
Solubility Plate Layout and SDS-PAGE Gel



These figures summarize the expression and solubility outcomes of the overall genome and specific categories. 1,451 proteins from *Shewanella Oneidensis* were processed in the method described above. These results can be further analyzed by the category of the protein. Both Conserved Hypotheticals and Periplasmic Proteins are shown.

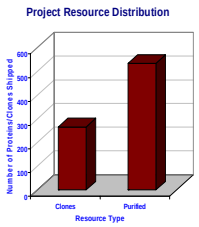


Periplasmic Vector, pBH31

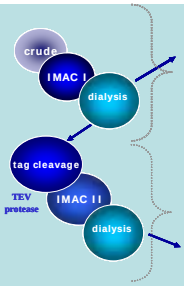


Expression and solubility results for 384 targets (4-96 well plates) directed to the cytoplasm or periplasmic compartment of *Escherichia coli*. Both systems provided expression efficiencies that averaged above 80% with approximately 50% of the proteins classified as soluble from both expression strategies. The combination of the two approaches provides positive expression outcomes of ~90% with greater than 60% of the expression products classified as soluble products. This study demonstrates the improved outcome for generation of soluble expression clones when multiple parallel expression are applied to a target group.

Resources: The project has successfully produced expression clones producing soluble protein as well as purified proteins. While we have cataloged ~800 soluble clones, we found an increased demand for the purified product. As seen in the graph to the right, over 250 clones have been distributed but we have observed a greater demand for purified proteins with a distribution of 532 purified samples. These resources in the form of validated expression clones and purified proteins are provided to collaborators and feedback is actively encouraged.



Protein purification strategy and results: Proteins were purified in parallel (4 per batch) using Qiagen Ni-NTA resin and a set of standard purification procedures that were developed for the high throughput structural genomics program. This process provides a proteins with a purity of 80-90%. After elution with imidazole, samples are dialyzed, concentrated, characterized and distributed to users. To improve purity , a secondary tag removal protocol cleaves the fusion component and results in further purification of the target. Cartoon illustrates the process flow for protein purification.



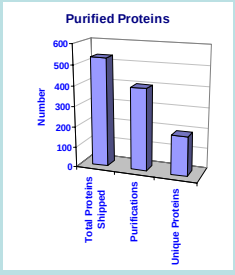
Post-purification

Buffer exchange
Concentration adjustment
Sequence validation/MS
UV/Vis CD scan
Aggregation status

Users

Functional Characterization
Protein Interactions
Affinity Tags
Protein chips
Mass Spec Standards

Purified proteins: The graph lists purified protein resources distributed to users. The total number of proteins shipped exceeds the number of purifications as some protein lots were sent to multiple users. The project has provided unique insight into the requirements and needs of the research community. For example, the capacity (both time and skill) needed for protein purification at a large scale may not be possible for many experimental projects. We have also noted that production scale and characterization requirements are diverse and dependent on the downstream use. This effort entails coordination of protein production with experimental requirements.



Domain Expression of simple membrane proteins: This domain cloning project used the SignalP and TMHMM programs to identify signal sequences and helical membrane spanning segments in *Geobacter sulfurreducens*. A set of rules was established to classify targets and generate confidence levels for proteins containing a single membrane spanning helix or membrane anchored proteins. This group of targets contains many two-component sensor and methyl accepting chemotaxis factor proteins. A parallel cloning approach was used to clone the same PCR fragments in both a cytoplasmic and periplasmic expression system. This approach will allow comparison of expression solubility outcomes for the same group of targets in two bacterial compartments and will provide a protein resource for characterization of the soluble domains and identification if interacting proteins.

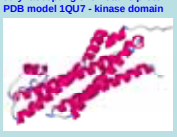
Simple Architecture Membrane Proteins

Purified Proteins

Structural models for the protein domains were obtained by submission of the protein sequence to the PHYRE (Protein Homology/analog Recognition Engine) server. Output from this analysis is a structural model with calculate E values for the aligned sequences as well as an estimated precision for the fold prediction. These data from fold recognition and the set of known structures were use as a general guide for comparison of the results obtained from the CD analysis

GSU1303 residues 31-184
methyl-accepting chemotaxis protein
PDB model 1OJG - sensory domain

GSU1303 residues 31-184
methyl-accepting chemotaxis protein
PDB model 1QU7 - kinase domain



This work was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, under contract W-31-109-Eng-38.

A U.S. Department of Energy laboratory managed by The University of Chicago